STRESYL

An Italian Stress-in-Syllables database for reading research

Simone Sulpizio, Giacomo Spinelli and Cristina Burani
Vita-Salute San Raffaele University, Milan/Fondazione Marica De Vincenzi onlus / University of Western Ontario / Institute of Cognitive Sciences and Technology/University of Trieste

During the last few decades, empirical research on reading has shown increasing interest in syllable units. More recently, stress assignment has become a particular focus of interest. The relation between syllables and stress, however, has yet to be investigated for Italian. In this paper, we describe a new database, STRESYL, that can help researchers to investigate the relation between syllables and stress in Italian. STRESYL offers type and token measures relating stress information to syllable units, both in terms of syllable forms and syllabic structures.

Keywords: syllable, syllabic structure, stress, database, Italian orthography

1. Introduction

For a long time, linguistic theories have identified the syllable as a core phonological unit of the linguistic system. The view that segments are organized into syllables, i.e., melodic sequences with rising and falling sonority, has been supported by converging evidence from a broad range of perspectives within phonology (e.g., Blevins 1995; Goldsmith 2011). In psycholinguistic and neurolinguistic research, interest in the syllable as a possible unit of processing has been increasing over the last few decades. Several studies have pointed out that syllables play an important role in word recognition and reading, being functional units of both visuo-orthographic decoding (e.g., Burani & Cafiero 1991; Carreiras, Alvarez & de Vega 1993; Ferrand & New 2003) and articulatory encoding (e.g., Cholin, Dell & Levelt 2011; Perret, Schneider, Dayer & Laganaro 2014).

Written Language & Literacy 20:1 (2017), **80–103.** DOI 10.1075/wll.20.1.05sul
ISSN 1387–6732 / E-ISSN 1570–6001

1.1 The role of syllables in reading

Most of the studies that have investigated the role of syllables in reading have focused on syllable frequency. In their seminal study, Carreiras et al. (1993) showed that syllable frequency affects word recognition in Spanish, with words beginning with a high-frequency syllable being recognized more slowly than words beginning with a low-frequency syllable. The same result has been replicated several times and in different languages (see, e.g., for French: Chetail, Colin & Content 2012; Mahé, Bonnefond & Doignon-Camus 2014; German: Conrad & Jacobs 2004). One explanation for the inhibitory effect of syllable frequency is the following: The initial syllable of a polysyllabic stimulus works as a functional reading unit and activates all the representations beginning with that syllable; in the case of a high-frequency syllable, lexical access will take longer because the correct target has to be selected among a large number of competitors (Conrad, Tamm, Carreiras & Jacobs 2010).

Syllable frequency also affects reading aloud and speech production. Across languages, several studies have shown that words and nonwords beginning with a high-frequency syllable are read aloud faster than words beginning with a low-frequency syllable (e.g., Carreiras, Mechelli & Price 2006; Laganaro & Alario 2006; Sulpizio & Job 2010, 2013). Syllables are posited to be articulatory programs that may be retrieved from a repertoire (when they are of high frequency) or assembled on-line (when they are of low frequency) on the basis of active phonological information. Thus, for a stimulus beginning with a high-frequency syllable, computation of the articulatory programs for production will be faster than for a stimulus beginning with a low-frequency syllable.

Few studies have investigated whether readers are affected by syllabic structure (i.e., the sequence of consonants and vowels (CV) making a syllable). In one study using the Stroop paradigm, Berent and Marom (2005) presented colored nonwords and asked participants to name the color of the stimuli in English. Importantly, the nonwords either shared or did not share their CV-structure with the color name (e.g., dult [CVCC] vs. dut [CVC] written in pink [CVCC]). The results showed that participants were faster to name the color when the nonword had a congruent CV-structure (e.g., *pof* written in red) than when the nonword had an incongruent CV-structure (e.g., *pof* written in green). Because participants in the Stroop task need to ignore the nonword distracter, this finding suggests that readers automatically (i.e., even when it is not mandatorily required by the task) assemble the phonological CV-structure of printed stimuli (see also Burani & Cafiero 1991; Marom & Berent 2010; for the activation of syllable structure from orthography, see, e.g., Chetail, Scaltritti & Content 2014; Perry, Ziegler & Zorzi 2010, 2014).

1.2 The marginalized relation between syllables and stress

Empirical investigations of syllabic units has generally been blind to the relation between syllables and stress (but see Sulpizio, Spinelli & Burani 2015). This blindness is also mirrored in existing lexical databases, which usually lack this type of information. However, the investigation of syllables in relation to stress may be highly fruitful for theories of reading.

Stress is the accentuation of a syllable within a word. In all languages that utilize stress, the syllables within polysyllabic words are not equal to each other, as some may be the locus of accentuation prominence, whereas others may not. Specifically, the stressed syllable is not only phonologically, but also acoustically more prominent than its unstressed counterpart. Thus, for example, in languages like English, Dutch or Italian, a stressed vowel is longer, louder, and with higher pitch than an unstressed one (Cutler 2005; Sulpizio & McQueen 2012).

The study of word recognition and reading can greatly benefit from investigating the representation of stressed and unstressed syllables, as such investigations may shed light on the processes of both stress assignment and syllable computation. Moreover, by studying the relation between stress and syllables, reading research may address more general issues, such as (a) how the reading system maps orthography to phonology and how it deals with opaque mappings (i.e., when there is no one-to-one correspondence, as in cases of an orthographic syllable that may be mapped onto either its phonologically stressed or unstressed counterpart), and (b) to what extent statistical regularities in orthography-to-phonology mappings affect the reading process (e.g., a reader might show a processing advantage in reading words bearing stress on a syllable (e.g., *ta*) that is usually stressed compared to words bearing stress on a syllable (e.g., *con*) that is usually unstressed).

Note that such reasoning applies not only to different syllable forms (e.g., *ta, ti, na*, etc.), but also to syllabic CV-structures (e.g., CV, CVC, CCV, etc.). In a given language, a syllabic structure (e.g., CV) might occur mainly as stressed. This might happen for distributional reasons, without involving any phonological rules of the language. In such cases, the distributional prominence of a given syllable might work as a bias within the reading system, which would be prone to associate stress with any syllable form possessing the same structure. However, the distribution of stress among different CV-structures might also be governed by some phonological rule. For example, the CV-structure of the penultimate syllable in a word is known to affect stress in Italian, with heavy syllables (i.e., those ending in a consonant) being more likely than light syllables (i.e., those not ending in a consonant) to bear stress (see, e.g., D'Imperio & Rosenthall 1999; but see Hayes (1995) for quantity-sensitive languages). However, further evidence is needed to deepen our understanding of the relation between CV-structure and stress. For instance, while Italian shows sensitivity to the weight of the penultimate syllable, whether the same is true for syllables in other word positions remains an open question. It might be the case that antepenultimate syllables attract stress more often when they are heavy than when they are light, just as penultimate syllables do. Alternatively, the CV-structure of syllables occurring in non-penultimate positions may have no role in driving stress assignment in Italian.

2. STRESYL

The study of the stress-syllable relation may particularly benefit from inspecting the statistical distributions that emerge from a frequency count of syllables where information concerning syllabic stressness is taken into account. By using such distributional information, researchers may be able to address empirical questions concerning the mechanisms and the processes of stress assignment and syllable computation and, more generally, the functioning of the reading system. To this aim, with the present paper, we offer a new tool, STRESYL (STREss-in-SYLlables); a database containing information about the stress patterns of both syllable forms and syllabic CV-structures in Italian. Information of this kind is neither present in the syllable frequency database for Italian by Stella and Job (2001), nor it is directly available in any current lexical databases for Italian (e.g., CoLFIS: Bertinetto, Burani, Laudanna, Marconi, Ratti, Rolando & Thornton, 2005; PhonItalia: Goslin, Galluzzi & Romani 2014), or for other free-stress languages (e.g., ELP: Balota, Yap, Hutchison, Cortese, Kessler, Loftis, Neely, Nelson, Simpson & Treiman 2007).

Italian can be of particular interest for its orthographic and phonological properties. First, in terms of its orthographic transparency and syllabic complexity Italian is close to Spanish, for which syllabic effects have been widely reported. In particular, Italian has an almost one-to-one correspondence between orthography and (segmental) phonology, although a few complex rules are required (Burani, Barca & Ellis 2006). With reference to syllabic structure, Italian is quite restrictive, especially regarding codas (Krämer 2009). Onset size ranges from three (e.g., STRAM.bo ‘strange’, capital letters indicate stressed syllable) to no segments (e.g., AM.bo ‘both’), whereas codas generally consist of a single consonant, which may be either the first part of a geminate (LET.to ‘bed’), a nasal (CAN.to ‘song’), or a liquid (TAL.pa ‘mole’, CAR.ta ‘paper’). Importantly, although there are fewer syllables ending in a vowel than syllables ending in a consonant in Italian (580 and 3,051, respectively), the former outweigh the latter in terms of the number of the words they appear in (70% vs. 30%) and the frequency of those words (72% vs. 28%; data extracted from PhonItalia). In part, this is due to the fact that no syllables ending in a consonant are allowed in a word-final position (most exceptions are loan words from English, such as *internet*). PhonItalia (Goslin et al. 2014) lists 3,626 unique syllables for Italian. Relative to languages such as English (which, according to some estimates, may have more than 12,000 unique syllables: Cholin et al. 2011), again, these figures are comparable with those obtained for Spanish (1,672 unique syllables were found using *BuscaPalabras*: Davis & Perea 2005).

A second aspect that makes Italian interesting for the present purposes is that it is a polysyllabic language with free-stress position: stress falls usually on one of the last three syllables, with a predominance of penultimate syllable stress. Specifically, about 80% of Italian words bear penultimate stress (e.g., maTIta, 'pencil'), whereas about 18% bear antepenultimate stress (e.g., TAvolo, 'table') (Thornton, Iacobini & Burani 1997). Moreover, stress assignment is not governed by rules and, in 98% of cases, it is not graphically marked.[1] Finally, less than 1% of Italian word forms are monosyllables (for which stress assignment is not an issue), whereas about 99% are polysyllables. Several studies addressing stress assignment in Italian have shown that readers are sensitive to the orthographic units within the stimulus, especially word endings (Burani & Arduino 2004; Burani, Paizi & Sulpizio 2014; Colombo 1992; Sulpizio, Arduino, Paizi & Burani 2013). However, research on syllables is almost absent from the psycholinguistic literature on Italian and much work has still to be conducted. Consider that the same orthographic syllable can be mapped onto different phonetic realizations depending on whether or not it bears stress (e.g., in Italian, *ta* → [ta] or [ta +stress], with the latter being realized as longer and louder than the former; note that the phenomenon occurs in all languages with free-stress position). Because stress information is not directly available in the syllable *per se*, an important empirical question is whether readers rely on the degree of association between the syllable and its stressed or unstressed realization.

STRESYL provides detailed information about the relation that exists between syllables and stress for each word form (the complete range of the database's fields is provided in the following section). Moreover, for each syllable form, the database includes the following information: (a) the number of different words containing the syllable (type numerosity information) for each stress pattern; (b) the summed frequency of words containing the syllable (token occurrences information) for each stress pattern. Such information is also provided for syllabic CV-structure. Note that, for each syllable form and each syllabic CV-structure, information is provided not only as an overall count – that is, irrespective of whether the syllable is the first or the last in the word – but also with reference to the

1. There is one rule to assign stress to three-syllable words that refers to the weight of the penultimate syllable: If it is heavy – that is, if it ends with a consonant (e.g., bi.SON.te, bison) – then the syllable attracts stress. However, there are different exceptions to the rule, for example, MAN.dor.la (almond) or LE.pan.to (Lepanto).

position the syllable occupies within the word. Thus, type and token counts are given for the same unit (e.g., *tra* in the case of the syllable form or, e.g., CCV in the case of syllabic CV-structure) when it occurs as the first, the penultimate, and the antepenultimate syllable of words (e.g., *tra* in tradizione, latrato, tranello, 'tradition, woof, trap'). The information concerning the first syllable is particularly relevant in reading, as syllable frequency effects have been reported mostly for the first syllable. Thus, the form of the first syllable may affect both the processes of lexical access and of stimulus articulation (e.g., Carreiras et al. 2006). The inclusion of counts concerning the penultimate and antepenultimate syllable within a polysyllabic word is driven by the Italian stress system, where the penultimate and the antepenultimate syllables bear stress in 98% of cases and there is no graphic marking of stress. With few exceptions, almost all the remaining 2% of words bear stress on the final syllable. In such cases, however, stress is graphically marked (e.g., coliBRÌ, 'hummingbird'), making stress assignment straightforward. Thus, the assignment of stress to the final syllable is not an issue for Italian readers, making that syllabic position totally different from the penultimate and antepenultimate positions; for this reason, we decided not to include counts concerning the last syllable in the database.

To sum up, the aim of the present work is to provide a database that provides information on the interplay between stress and syllables that is currently not available for Italian. In so doing, STRESYL may constitute a useful and effective tool for psycholinguistic and neurolinguistic research interested in word recognition and reading.

2.1 The database

STRESYL has been constructed on the basis of information contained in PhonItalia 1.1 (Goslin et al. 2014; http://phonitalia.org), the most recent database for the Italian lexicon that contains information about word stress.

STRESYL comes in two components. The first component is an augmented version of PhonItalia. All word forms included in PhonItalia 1.1 (120,000 in total) were divided into syllables by means of the Hyphen library of the Hunspell spell checker (Németh 2006). Word forms were syllabified according to the standard orthographic rules for Italian.[2] Note that orthographic syllabification differs

2. Note that, according to the standard syllabification for Italian orthography, syllabic boundaries fall between geminate consonants (e.g., MAS.sa 'mass'). In order to be consistent with the syllabification used for syllable forms, CV-structures treated geminate consonants simply as double consonants (i.e., CC) with a syllabic boundary between them (e.g., the syllabified CV-structure for *massa* is CVC.CV).

from the phonological syllabification of standard Italian pronunciations (which is provided by PhonItalia) in a limited number of cases, such as preconsonantal /s/ and clusters mapping to geminate consonants (gn, gli).[3] Our aim was to provide a database for research addressing orthographic cues to stress assignment, in line with the literature (e.g., Jouravlev & Lupker 2014; Ševa, Monaghan & Arciuli 2009; Sulpizio et al. 2013). Thus, we focused on orthographic syllable boundaries (cf. Stella & Job 2001). Syllable boundaries for each word form were projected onto the corresponding orthographic CV-structure in order to yield CV-structures for each syllable within each word. In this way we could add nine new fields to PhonItalia.[4]

Table 1 presents the new fields added to Phonitalia 1.1. In the first three new fields, we report the syllable forms appearing in the first, antepenultimate, and penultimate positions for each word form. When a word form had no syllable in the penultimate or antepenultimate positions (because it was monosyllabic or disyllabic, respectively), the corresponding fields received a N/A (i.e., 'not available') value. Fields 4 to 6 report syllabic CV-structures in the same way. Finally, in fields 7 to 9, each of the aforementioned syllable positions received three different values according to whether the corresponding syllable was stressed (1), unstressed (0), or not available (N/A). Thus, data for the first three syllables are reported for four-syllable words, such as *solletico* 'tickle', because they correspond to the first, the antepenultimate and the penultimate syllable positions, respectively (Table 1, first row). In contrast, for three-syllable words, such as *ombrello* 'umbrella', only data for two syllables are available (Table 1, second row), whereas for disyllabic and monosyllabic words, such as *cane* 'dog' and *re* 'king', only data for one syllable are available (Table 1, third and fourth rows). Note, futher, that the syllables in the first and the penultimate positions coincide within disyllabic words, as do the syllables in the first and the antepenultimate positions of three-syllable words. This

3. Regional Italian differs in this respect, with clusters such as gn and gl being pronounced as single consonants in Northern Italy.

4. Sometimes, stress marks are used in syllables other than final (e.g., when two homographs with different stress position have to be disambiguated). These cases were excluded from analyses and tables; however, they were reported in the database. In PhonItalia (i.e., the source of our database), stress marks are represented by means of apostrophes. However, PhonItalia does not specify a representation for stress diacritics in CV-structures, nor does it distinguish cases in which apostrophes stand for stress marks (e.g., *citta'* 'city') from those in which they are actual apostrophes (e.g., *D'Annunzio* 'D'Annunzio'). As this type of diacritics is relevant to stress assignment, CV-structures obtained from PhonItalia were slightly modified by transcribing apostrophes with two special marks. Vowels followed by apostrophes standing for stress marks were transcribed with an S (Stressed vowel); the remaining apostrophes were transcribed with an A (Apostrophe).

Table 1. The nine new fields added to PhonItalia 1.1

Old fields	New fields								
	Syllable forms			Syllabic CV-structures			Syllabic stressness		
	1	2	3	4	5	6	7	8	9
Word form	First syllable	Antepenult syllable	Penultimate syllable	First syllable	Antepenult syllable	Penultimate syllable	First syllable	Antepenult syllable	Penultimate syllable
solletico	sol	le	ti	CVC	CV	CV	0	1	0
ombrello	om	om	brel	VC	VC	CCVC	0	0	1
cane	ca	N/A	ca	CV	N/A	CV	1	N/A	1
re	re	N/A	N/A	CV	N/A	N/A	1	N/A	N/A

Note. For each word form in PhonItalia 1.1, the new fields report data concerning syllables in the first, the antepenultimate and the penultimate positions (First syllable, Antepenult syllable, Penultimate syllable). Syllable forms are reported in fields 1 to 3, syllabic CV-structures in fields 4 to 6, and syllabic stressness (with 0 = unstressed, 1 = stressed, N/A: not available) in fields 7 to 9.

component of STRESYL can be downloaded from the following link: http://www.istc.cnr.it/grouppage/databases

The second component of STRESYL includes Summary Tables (see, STRESYL_TY.xls and STRESYL_TK.xls; http://www.istc.cnr.it/grouppage/databases), a set of charts listing all syllable forms and syllabic CV-structures. For each syllable form and syllabic CV-structure, there is indication of how many times it appears as either stressed or unstressed, as both number of occurrences (e.g., 300 times stressed vs 1,500 times stressed) and percentages (e.g., 20% of times stressed vs 80% of times unstressed). Syllable units in each sheet are sorted by total occurrences (i.e., how many times the syllable occurs regardless of stress) in descending order. All measures are available as both overall and positional counts, and as both types (STRESYL TY.xls) and tokens (STRESYL TK.xls), as described below (see Tables 3a and 3b for examples). The Summary Tables constitute a remarkable tool for research, as they allow a researcher to browse through and detect syllable units with varying distributional characteristics with respect to stress. With this tool, researchers will be able to search for both specific syllable forms and syllabic CV-structures. Syllable forms and syllabic CV-structures can then be used to query the augmented version of PhonItalia (i.e., the first component of our database) in order to extract all the word forms in which they appear in any of the available positions.

2.2 Type and token measures

The type frequency measure refers to the number of different words that contain the specific unit. For example, the type frequency of the syllable *pa* indicates the number of words that contain this syllable (e.g., patata, panino, riparo, etc. 'potato, sandwich, refuge'). In STRESYL, type measures were obtained by counting how many times a given syllable form or syllabic CV-structure was stressed or unstressed in a given position. Thus, counts for each syllable position (except the last one) were collapsed in order to have overall stressness measures for each syllable. The last syllable position was excluded from the overall measure because, as stated above, words with final stress in Italian are few and always graphically marked.[5]

5. PhonItalia lists a single word form two or more times if it appears in both syntagmatic and non-syntagmatic occurrences (e.g., mangiato 'eaten' with and without a preceding auxiliary verb). We felt that this distinction would lead to overrating stress type counts for some syllables (especially those occurring in the penultimate position of past participle forms, such as mangiato). Thus, we counted cases of this sort just once. In the token measures, however, when a single word form was listed twice or more, the frequency of each entry was considered and summed up.

The token frequency measure, instead, refers to the summed frequency of all the words that contain the specific unit (e.g., if the occurrences of *pa* were restricted to the examples cited above, then its token frequency would be 91, i.e., 25 (frequency for *patata*) + 15 (frequency for *panino*) + 51 (frequency for *riparo*)).

In the literature, there is debate over the general impact of type and token frequencies on lexical processing (e.g., Clahsen 1999; del Prado Martín, Ernestus & Baayen 2004). The relevance of each measure may depend on the variable of interest and the experimental paradigm adopted. For instance, it has been shown that token frequency is responsible for the inhibitory effect of syllable frequency on lexical decision (Conrad, Carreiras & Jacobs 2008), whereas it is not clear whether the facilitatory effect that syllable frequency exerts on reading aloud is due to type or token frequency. In contrast, type frequency seems to be more important than token frequency within the literature on stress assignment. Results from both word reading (Burani & Arduino 2004) and nonword reading (Sulpizio et al. 2013) suggest that the more word types that share a given orthographic ending (e.g., -*ola*), the more the ending is effective in driving the assignment of the stress pattern associated with the majority of those word types (e.g., most words ending in -*ola* bear antepenultimate stress: PENt*ola*, TOMb*ola*, BAMb*ola*, etc. 'pot, bingo, doll'). The token frequency of words sharing an ending, however, does not seem to matter in the neighborhood (Sulpizio et al. 2013).

Considering the above results, both type and token frequency measures may be of interest for research on syllables. By providing both measures, STRESYL allows researchers to investigate which of them impacts more on word recognition and reading as far as syllabic stressness is considered.

3. Statistics

3.1 Descriptive analyses

Tables 2a and 2b show the descriptive statistics for the variables included in STRESYL. Statistics are reported for both syllable forms and syllabic CV-structures, and for both overall counts and each syllable position. Tables 3 and 4 provide two examples of the Summary Tables, i.e., tables listing how many times the different syllable forms and CV-structures occur as stressed or unstressed (see above).

Syllable forms

A first look at the means for the positional counts (Table 2a) shows that the numerosity and the frequencies of stressed and unstressed syllables in the different positions reflect the distributional properties of the Italian stress system: When

Table 2a. Descriptive statistics for the measures included in the database for the syllable units. Standard deviations are reported in parenthesis

Variables	Overall			First syllable			Antepenultimate syllable			Penultimate syllable		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Stressed Types – absolute nr.	41.52 (181.04)	0	3317	8.43 (18.45)	0	255	9.93 (33.65)	0	461	41.14 (170.66)	0	2854
Unstressed Types – absolute nr.	77.76 (328.38)	0	5783	49.47 (221.89)	0	4464	41.41 (129.72)	0	1485	11.58 (75.96)	0	1539
Stressed Types – percentage	47.8 (36.33)	0	100	37.25 (34.59)	0	100	22.44 (31.31)	0	100	89.55 (22.05)	0	100
Unstressed Types – percentage	52.2 (36.33)	0	100	62.75 (34.59)	0	100	77.56 (31.31)	0	100	10.45 (22.05)	0	100
Stressed Tokens – absolute nr.	848.23 (3428.4)	0	44721	566.26 (2456.51)	0	38548	111.5 (463.29)	0	8184	926.76 (3477.13)	0	41627
Unstressed Tokens – absolute nr.	792.99 (3767.9)	0	63375	603.13 (2984.64)	0	57735	533.04 (2056.37)	0	37584	137.96 (875.96)	0	16606
Stressed Tokens – percentage	55.61 (37.92)	0	100	48.87 (38.1)	0	100	22.01 (33.38)	0	100	91.31 (22.6)	0	100
Unstressed Tokens – percentage	44.38 (37.92)	0	100	51.13 (38.1)	0	100	77.99 (33.38)	0	100	8.69 (22.6)	0	100

Note. The statistics are expressed as both mean values calculated on the absolute number of occurrences (out of a total of 108,974) and as percentages of cases for the total occurrences of types and tokens, respectively (i.e., the percentage for stressed types is calculated from the total types, etc.). The same statistics are reported for the overall counts (Overall) and for each positional count (First syllable, Antepenultimate syllable, Penultimate syllable).

Table 2b. Descriptive statistics for the measures included in the database for the syllabic CV-structures. Standard deviations are reported in parenthesis

Variables	Overall			First syllable			Antepenultimate syllable			Penultimate syllable		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Stressed Types – absolute nr.	4122.24 (10712.35)	0	48117	629.36 (1310.8)	0	4793	740.44 (1923.48)	0	8748	3369.36 (8765.88)	0	39210
Unstressed Types – absolute nr.	7720.92 (20078.04)	0	95111	3685.16 (7780.02)	0	31376	3083.24 (8133.9)	0	37882	948.52 (3680.11)	0	18447
Stressed Types – percentage	38.85 (27.98)	0	100	28.95 (27.08)	0	100	28.76 (28.95)	0	100	86.27 (20.87)	0	100
Unstressed Types – percentage	61.15 (27.98)	0	100	71.05 (27.08)	0	100	71.24 (28.95)	0	100	13.73 (20.87)	0	100
Stressed Tokens – absolute nr.	84212.48 (200598.1)	0	881478	42176.72 (88294.71)	0	342806	8301.28 (20751.52)	0	95902	75848.88 (180061.5)	0	784794
Unstressed Tokens – absolute nr.	78728.72 (206848.6)	0	983579	44921.56 (99146.23)	0	411234	39679.12 (88294.71)	0	473131	11290.92 (43253.62)	0	216948
Stressed Tokens – percentage	50.73 (28.83)	0	100	52.32 (27.68)	0	100	27.8 (30.15)	0	100	89.25 (21.87)	0	100
Unstressed Tokens – percentage	49.27 (28.83)	0	100	47.68 (27.68)	0	100	72.2 (30.15)	0	100	10.75 (21.87)	0	100

Note. The statistics are expressed as both mean values calculated on the absolute number of occurrences (out of a total of 108,974) and as percentages of cases for the total occurrences of types and tokens, respectively (i.e., the percentage for stressed types is calculated from the total types, etc.). The same statistics are reported for the overall counts (Overall) and for each positional count (First syllable, Antepenultimate syllable, Penultimate syllable).

occurring in the penultimate position, syllables bear stress ~90% of the time. Moreover, the mean token frequency of words where syllables appear in the penultimate position is higher when they are stressed than when they are unstressed (926.76 and 137.96, respectively). For the first and the antepenultimate positions, the pattern is reversed. However, if one looks at the overall counts, both stressed and unstressed types and stressed and unstressed tokens tend to have a similar distribution (i.e., around 50%). This fact suggests that syllable forms are not a priori biased toward either stressness or unstressness (but see below).

Beyond considering average values, it is also important to look at the Summary Tables that show how individual syllable forms behave in terms of numerosity and frequency (see Table 3 for a sample of the Summary Table for syllable forms in the penultimate position (type measure); see STRESYL_TY.xls, sheets 1–4, and STRESYL_TK.xls, sheets 1–4 for the complete Summary Tables of syllable forms). Inspection of single cases in the Summary Tables reveals that, for positional counts, the types and tokens measures for each syllable form mirror those for the average values, although some cases slip away from the main tendency. For example, when all syllables in the penultimate position are considered, most show a majority of stressed types and tokens, but there are some cases (~6% of syllables) that exhibit the opposite pattern (for some examples, look up *bi, do*, and *te* in Table 3). The same is true for the first and antepenultimate positions: These include a majority of syllables with a higher number of unstressed types, but some syllables show the opposite trend with a majority of stressed types (first position: ~25%, e.g.: *bul, es, dra*; antepenultimate position: ~17%, e.g., *bi, reb, tan*). Note that, from a statistical learning point of view, the infrequent cases that exhibit an antidominant trend are especially interesting, as they may be critical for testing the sensitivity of readers to the distributional properties of syllabic stressness (see Section 4 below).

Results for overall counts provide an interesting insight on the average balance of stressed and unstressed syllable forms reported above. Rather than reflecting the general behavior of individual syllable forms, that finding seems to be a byproduct of two opposite groups of syllables, one with a majority of stressed types (~40% of syllables bear stress in >55% of their types) and one with a majority of unstressed types (~52% of syllables do not bear stress in >55% of their types).[6] Note that the same reasoning also applies to the token counts.

6. The remaining ~8% are syllables with no bias towards either stressness or unstressness (i.e., the proportions of their stressed and unstressed types were between 45% and 55%).

Table 3. Sample of the Summary Table for syllable forms in the penultimate position (type measure)

Syllable form	Stressed types – absolute nr.	Unstressed types – absolute nr.	Total types – absolute nr.	Stressed types – percentage	Unstressed types – percentage
ta	2854	374	3228	0.88	0.12
ti	1948	1239	3187	0.61	0.39
na	2428	185	2613	0.93	0.07
ra	2070	190	2260	0.92	0.08
men	2142	14	2156	0.99	0.01
to	1777	309	2086	0.85	0.15
ca	1804	204	2008	0.90	0.10
la	1770	165	1935	0.91	0.09
ri	1437	483	1920	0.75	0.25
va	838	1075	1913	0.44	0.56
li	1292	606	1898	0.68	0.32
zio	1882	10	1892	0.99	0.01
do	258	1539	1797	0.14	0.86
ni	876	658	1534	0.57	0.43
si	765	625	1390	0.55	0.45
bi	249	1113	1362	0.18	0.82
te	469	871	1340	0.35	0.65
za	1184	90	1274	0.93	0.07
di	757	430	1187	0.64	0.36
ma	965	152	1117	0.86	0.14
re	862	216	1078	0.80	0.20
da	906	167	1073	0.84	0.16
sa	932	105	1037	0.90	0.10

Note. For each syllable form, the counts indicate how many times it appears as either stressed or unstressed in the penultimate position. Data are sorted by total frequency in descending order. The statistics are expressed both in terms of the absolute number of occurrences out of a total of 108,974 occurrences and as the percentage of cases out of the total occurrences for each syllable form.

Syllabic CV-structure

The descriptive statistics for syllabic CV-structure (Table 2b) show distributions that are similar to those for syllable forms, both for positional and overall counts. Looking at the means for the positional counts (Table 2b), one can see that numerosity for stressed and unstressed syllables in the different positions reflects the distribution of Italian stress patterns: syllables in the penultimate position occur more in words in which they receive stress; in the first and antepenultimate positions, the pattern is reversed with syllables occurring more often in words in which they are unstressed. The overall count shows a slight advantage for unstressed over stressed types (61% vs. 39%), which suggests that syllabic CV-structures have a slight bias toward unstressness. This might reflect that most of the Italian lexicon is composed of polysyllabic words with at least three syllables and in these cases, for each stressed syllable, there are at least two syllables without stress.

With regard to tokens, while the patterns for the antepenultimate and penultimate positions parallel those reported for types, both the first position and the overall count show a similar proportion of stressed and unstressed tokens (first position: 52% vs. 48%; overall count: 51% vs. 49%). The last result suggests that, in terms of frequency of occurrence, syllabic CV-structures do not show any stress bias.

When we look at the Summary Tables, inspection of the different types of syllabic CV-structures for the overall count shows that, in terms of types, most of the structures have a tendency toward being either stressed or unstressed (see Table 4 for the Summary Table of syllabic CV-structures in any position (type measure); see STRESYL_TY.xls, sheets 5–8, and STRESYL_TK.xls, sheets 5–8 for all Summary Tables for syllabic CV-structures). In particular, a few structures have a strong tendency toward unstressness (e.g., V and VC are unstressed 81% and 93% of times, respectively), but none show a strong tendency toward stressness. Tokens show a more heterogeneous distribution, with at least some cases representing the different possibilities (high stress dominance, low stress dominance, high unstress dominance, low unstress dominance).

Concerning the positional counts, we note that for penultimate positions, almost all structures have a prevalence of stressed types and tokens. Furthermore, structures ending in a consonant bear stress relatively more often than structures ending in a vowel; a finding that is supportive of the view that the weight of the penultimate syllable affects the position of stress in Italian (D'Imperio & Rosenthall 1999). The pattern is reversed for the antepenultimate position, which shows a prevalence for unstressed types and tokens. Again, the data are in line with the distribution of Italian stress patterns. The first position counts offer a rather different pattern as the majority of structures have a larger number of unstressed types, but the proportions are almost the same for stressed and unstressed tokens (48% vs. 52%).

Table 4. Summary Table for overall syllabic CV-structures (type measure)

Syllabic CV-structure	Stressed types – absolute nr.	Unstressed types – absolute nr.	Total types – absolute nr.	Stressed types – percentage	Unstressed types – percentage
CV	48117	95111	143228	33.59	66.41
CVC	26807	37486	64293	41.70	58.30
CCV	6592	16198	22790	28.92	71.08
VC	1313	16835	18148	7.23	92.77
CVV	8478	5604	14082	60.20	39.80
V	2184	9659	11843	18.44	81.56
CCVC	4080	6083	10163	40.15	59.85
CVVC	1673	1452	3125	53.54	46.46
CCVV	1687	1264	2951	57.17	42.83
CCCV	567	1353	1920	29.53	70.47
CCCVC	478	625	1103	43.34	56.66
VV	274	711	985	27.82	72.18
CCVVC	558	358	916	60.92	39.08
CCCVV	106	89	195	54.36	45.64
CCCVVC	43	58	101	42.57	57.43
VVC	45	52	97	46.39	53.61
CVVV	38	24	62	61.29	38.71
CVCC	5	53	58	8.62	91.38
CCVCC	5	4	9	55.56	44.44
VVV	3	0	3	100.00	0.00
CCVVV	3	0	3	100.00	0.00
VCCC	0	1	1	0.00	100.00

Note. For each CV-structure, the counts indicate how many times it appears as either stressed or unstressed in the penultimate position. Data are sorted by total frequency in descending order. The statistics are expressed both in terms of the absolute number of occurrences out of a total of 108,974 occurrences and as the percentage of cases out of the total occurrences of each CV-structure.

3.2 Correlation analyses

In order to explore the relation between our measures, we ran correlation analyses. Such analyses help to highlight whether there is any relation among the measures we offer. We focused on the correlation between the stressed and unstressed counts for the same measure in each position (e.g., stressed and unstressed type in the penultimate position), as this correlation allows us to infer whether a particular syllable distributes similarly when it occurs as stressed or unstressed in any position or in a particular position.

Before running the analyses, we checked whether the measures were normally distributed. As that is not the case (i.e., all distributions are leptokurtic and highly right-skewed), Spearman correlations were run. Tables 5a and 5b report the correlation matrixes for the analysis of syllable forms and syllabic CV-structure,

Table 5a. Correlation matrix for the variables of syllable forms (*** indicate $p < .001$)

Variables	Stressed Types	Stressed Tokens
Overall		
Unstressed Types	.64***	–
Unstressed Tokens	–	.60***
First syllable		
Unstressed Types	.63***	–
Unstressed Tokens	–	.57***
Penultimate syllable		
Unstressed Types	.49***	–
Unstressed Tokens	–	.46***
Antepenultimate syllable		
Unstressed Types	.55***	–
Unstressed Tokens	–	.51***

Table 5b. Correlation matrix for the variables of syllabic CV-structure (*** indicate $p < .001$)

Variables	Stressed Types	Stressed Tokens
Overall		
Unstressed Types	.94***	–
Unstressed Tokens	–	.96***
First syllable		
Unstressed Types	.95***	–
Unstressed Tokens	–	.97***
Penultimate syllable		
Unstressed Types	.95***	–
Unstressed Tokens	–	.94***
Antepenultimate syllable		
Unstressed Types	.96***	–
Unstressed Tokens	–	.97***

respectively. Note that, for the syllable forms, the samples used for the analyses were very large and caution is needed in interpreting the results. With large sample sizes, the chance of obtaining significant p-values is high even when the correlation coefficient is very small. Note, however, that as the strength of the relation between the explored measures is given by the correlation coefficient, that information should be considered, more than significance.

Syllable forms

For the overall count, we found high correlations between stressed and unstressed types, and between stressed and unstressed tokens. This result suggests that stressed and unstressed syllables show similar distributions: Syllables appearing in few (and infrequent) unstressed words, also appear in few (and infrequent) stressed words, whereas syllables appearing in many (and frequent) unstressed words also appear in many (and frequent) stressed ones.

The correlation analyses for positional syllables – that is, for the first, antepenultimate, and penultimate positions – reveal somewhat lower values, but show a very similar pattern to that reported for overall counts.

Syllabic CV-structure

The correlation analyses yield results similar to those for syllable forms, with even higher correlation coefficients, suggesting that, in terms of frequency distributions, all the measures behave in almost the same way. For the overall counts, stressed and unstressed types are highly correlated with each other, as well as stressed and unstressed tokens. Exactly the same pattern emerges in all analyses for positional syllables.

4. Summary and future directions

Here we have described STRESYL, a new tool for psycholinguistic and neurolinguistic research interested in the study of syllables in relation to stress in Italian. To our knowledge, the issue has been almost totally neglected within the literature, but it has great relevance for theories of word recognition and reading, as well as computational models that aim to simulate performance for the two tasks. Without claiming to be exhaustive, we highlight below some of the most relevant data that emerges from querying the database and the issues such data pose for researchers.

Relevant information that emerges from the results of the correlation analyses is that syllable forms and syllabic-CV-structures exhibit similar behavior in terms of occurring in either their stressed or unstressed versions. Thus, a syllable occurring in many words as unstressed will also occur in many other words as stressed, suggesting that the syllable's behavior is not affected by stress information and that there is no *a priori* bias toward either stressness or unstressness. This is also suggested by the descriptive statistics for the overall syllable counts, as both stressed and unstressed syllable forms occur approximately 50% of the time.

When looking at the positional counts, however, one observes a bias in accordance with the stress distribution of the language: the penultimate position shows a dominance for stressed cases and the antepenultimate (and first) position shows a dominance for unstressed ones. Note, however, that in a considerable number of cases, the stress distribution for syllable forms is reversed. Thus, for example, ~17% of syllable forms in the antepenultimate position occur mainly as stressed: some of them have a middle-large number of types (e.g., *tan*); others, although having a very small number of types, show a very high bias, with some cases occurring rarely but always stressed (e.g., *tis* has 30 stressed types out of a total of 30 types). These cases are of particular interest as they might work as stress cues for the reading system, being more effective than the overall bias in driving participants to assign stress.

With regard to the descriptive analyses of syllabic CV-structure, the positional counts show a pattern analogous to that already described for syllable forms. The overall counts, instead, show a different trend, with a slight bias toward unstressed types (61% vs. 39%). A significant contribution to this bias comes from some structures, such as V, VC, and CVCC, which have a marked tendency to being unstressed (81%, 93%, and 91% of times, respectively). Because of their strong biases, these structures are of particular interest for empirical research, which may investigate to what extent such biased structures affect stress assignment during word reading.

The data emerging from STRESYL, as well as the types of information contained in it will allow researchers to empirically investigate several issues pertaining to the representation of syllables within the reading system and their interaction with other phonological information, such as stress. The first relevant issue that STRESYL may help to investigate is the effect of syllable frequency in relation to syllabic stressness. Previous research has shown that syllable frequency affects word recognition and reading. The effect is driven by token frequency and has a phonological nature in that it involves phonological syllables (Conrad et al. 2008; Conrad, Grainger & Jacobs 2007). No one, however, has investigated yet whether syllable frequency effects occur only when considering stressed or unstressed syllables, or irrespective of syllabic stressness. In the case of reading aloud, a high-frequency syllable speeds up the retrieval of motor programs of the to-be-planned unit, which is supposed to be stored in a syllabic repertoire (e.g., Cholin et al. 2011). But, are such stored units represented as stressed or unstressed? Does this depend on their overall or positional tendency? Answering these questions would be informative about the nature of syllable representations and, more generally, about the relation between syllabic and stress information.

A second issue for theories of reading concerns the relation that exists between stress information driven by syllables that occur in a relevant position within a word, like the first and the penultimate ones, and other sources of information known to affect stress assignment, such as lexical stress and/or orthographic sequences (for a different and complementary tool for the investigation of stress in reading Italian, see Spinelli, Sulpizio & Burani 2016). For example, in many cases lexical stress and syllabic stressness converge, but what happens when these two types of information diverge? This is the case, for example, for the word *biDOne* (bin), which is stressed on the penultimate syllable, while the syllable *do* is usually unstressed when occurring in the penultimate position. Does this stress contrast produce any costs for the system? It is worth noting that the same issue applies to syllabic CV-structures, where the same questions deserve investigation.

The relation between stress and syllables may also have relevant implications for word recognition. Recently, it has been shown that, in Italian, lexical stress may affect the recognition process, such that participants are slower when recognizing antepenultimate compared to penultimate stress words (Colombo & Sulpizio 2015). As syllables are relevant units for word recognition, it can be asked whether syllable and stress may interact during word recognition, and what is the dynamics between lexical stress and syllable stressness (i.e., the fact that a syllable may appear mostly as either stressed or unstressed).

Finally, we believe that, although STRESYL is primarily a research tool, it can also become a useful support for teachers. In a language with a transparent orthography like Italian, syllables play an important role in the teaching of reading practice (see, e.g., De Coster, Baidak, Motiejunaite & Noorani 2011; for software that supports reading, see, e.g., Vio & Tretti 2008). If we consider that, in Italian, the relation between syllable forms and stress is a case of orthography-to-phonology ambiguity and stress assignment is one of the main difficulties for young readers – second graders make stress errors but almost no segmental errors (Sulpizio & Colombo 2013) – STRESYL might help teachers to promote phonological awareness by focusing on the fact that orthographic syllables are phonetically ambiguous. In fact, the same orthographic syllable may occur as either stressed or unstressed and, therefore, has different phonetic realizations. The same could be done by teachers of Italian as L2, since stress assignment is also an issue for L2 learners, both children (Bellocchi, Bonifacci & Burani 2016) and adults (Primativo, Rinaldi, O'Brien, Paizi, Arduino & Burani 2013).

The issues that we have sketched above are intended to be only a few of the examples that STRESYL may help to investigate. STRESYL can be particularly helpful for advancing reading theories as it allows researchers to address multiple relevant questions, in different types of populations (e.g., adults and children; typical and atypical developing readers), and with different types of stimuli (e.g., words and nonwords): In so doing, STRESYL will help to advance our knowledge of word recognition and reading aloud.

References

Balota, David A., Melvin J. Yap, Keith A. Hutchison, Michael J. Cortese, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson, & Rebecca Treiman. (2007). The English lexicon project. *Behavior Research Methods* 39: 445–459.

Bellocchi, Stephanie, Paola Bonifacci & Cristina Burani. (2016). Lexicality, frequency and stress assignment effects in bilingual children reading Italian as a second language. *Bilingualism: Language and Cognition* 19: 89–105.

Berent, Iris & Michal Marom. (2005). Skeletal structure of printed words: Evidence from the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance* 31: 328–338.

Bertinetto, Pier Marco, Cristina Burani, Alessandro Laudanna, Lucia Marconi, Daniela Ratti, Claudia Rolando & Anna M. Thornton. (2005). Corpus e Lessico di frequenza dell'italiano scritto (CoLFIS) [CoLFIS. Corpus and frequency lexicon of written Italian]. Retrieved from http://www.istc.cnr.it/material/database/ colfis/index_eng.shtml

Blevins, Juliette. (1995). The syllable in phonological theory. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 206–244. Oxford: Blackwell.

Burani, Cristina & Lisa S. Arduino. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain and Language* 90: 318–325.

Burani, Cristina, Laura Barca & Andrew W. Ellis. (2006). Orthographic complexity and word naming in Italian: Some words are more transparent than others. *Psychonomic Bulletin & Review* 13: 346–352.

Burani, Cristina & Rosaria Cafiero. (1991). The role of sub-syllabic structure in lexical access to printed words. *Psychological Research* 53: 42–52.

Burani, Cristina, Despina Paizi & Simone Sulpizio. (2014). Stress assignment in reading Italian: Friendship outweighs dominance. *Memory & Cognition* 42: 662–675.

Carreiras, Manuel, Carlos J. Alvarez & Manuel De Vega. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language* 32: 766–780.

Carreiras, Manuel, Andrea Mechelli & Cathy J. Price. (2006). Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Human brain mapping* 27: 963–972.

Chetail, Fabienne, Cécile Colin & Alain Content. (2012). Electrophysiological markers of syllable frequency during written word recognition in French. *Neuropsychologia* 50: 3429–3439.

Chetail, Fabienne, Michele Scaltritti & Alain Content. (2014). Effect of the consonant – vowel structure of written words in Italian. *The Quarterly Journal of Experimental Psychology* 67: 833–842.

Cholin, Joana, Gary S. Dell & Willem JM Levelt. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37: 109–122.

Clashen, Harald. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral & Brain Sciences* 22: 991–1060.

Colombo, Lucia. (1992). Lexical stress effect and its interaction with frequency in word pronunciation. *Journal of Experimental Psychology: Human Perception and Performance* 18: 987–1003.

Colombo, Lucia & Simone Sulpizio. (2015). When orthography is not enough: The effect of lexical stress in lexical decision. *Memory & Cognition* 43: 811–824.

Conrad, Markus, Manuel Carreiras & Arthur M. Jacobs. (2008). Contrasting effects of token and type syllable frequency in lexical decision. *Language and Cognitive Processes* 23: 296–326.

Conrad, Markus, Jonathan Grainger & Arthur M. Jacobs. (2007). Phonology as the source of syllable frequency effects in visual word recognition: Evidence from French. *Memory & Cognition* 35: 974–983.

Conrad, Markus & Arthur M. Jacobs. (2004). Replicating syllable frequency effects in Spanish in German: One more challenge to computational models of visual word recognition. *Language and Cognitive Processes* 19: 369–390.

Conrad, Markus, Sascha Tamm, Manuel Carreiras & Arthur M. Jacobs. (2010). Simulating syllable frequency effects within an interactive activation framework. *European Journal of Cognitive Psychology* 22: 861–893.

Cutler, Anne. (2005). Lexical Stress. In David B. Pisoni & Robert E. Remez (eds.), *The handbook of speech perception*, 264–289. Oxford: Blackwell.

Davis, Colin J. & Manuel Perea. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods* 37: 665–671.

De Coster, Isabelle, Nathalie Baidak, Akvile Motiejunaite & Sogol Noorani. (2011). *Teaching Reading in Europe: Contexts, Policies and Practices*. Education, Audiovisual and Culture Executive Agency, European Commission. Available from EU Bookshop.

del Prado, Martín, Fermín Moscoso, Mirjam Ernestus & R. Harald Baayen. (2004). Do type and token effects reflect different mechanisms? Connectionist modeling of Dutch past-tense formation and final devoicing. *Brain and language* 90: 287–298.

D'Imperio, Mariapaola & Sam Rosenthall. (1999). Phonetics and phonology of main stress in Italian. *Phonology* 16: 1–28.

Ferrand, Ludovic & Boris New. (2003). Syllabic length effects in visual word recognition and naming. *Acta Psychologica* 113: 167–183.

Goldsmith, John. (2011). The syllable. In John Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The handbook of phonological theory*, 164–196. Oxford: Blackwell.

Goslin, Jeremy, Claudia Galluzzi & Cristina Romani. (2014). PhonItalia: a phonological lexicon for Italian. *Behavior research methods* 46: 872–886. Retrieved from http://phonitalia.org/

Hayes, Bruce. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.

Jouravlev, Olessia & Stephen J. Lupker. (2014). Stress consistency and stress regularity effects in Russian. *Language, Cognition and Neuroscience* 29: 605–619.

Krämer, Martin. (2009). *The phonology of Italian*. Oxford: Oxford University Press.

Laganaro, Marina & F.-Xavier Alario. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language* 55: 178–196.

Mahé, Gwendoline, Anne Bonnefond & Nadège Doignon-Camus. (2014). The time course of the syllable frequency effect in visual word recognition: Evidence for both facilitatory and inhibitory effects in French. *Reading and Writing* 27: 171–187.

Marom, Michal & Iris Berent. (2010). "Phonological constraints on the assembly of skeletal structure in reading. *Journal of psycholinguistic research* 39: 67–88.

Németh, László. (2006). Automatic non-standard hyphenation in OpenOffice.org. *TUGboat*, 27, 32–37.

Perret, Cyril, Laurence Schneider, Géraldine Dayer & Marina Laganaro. (2014). Convergences and divergences between neurolinguistic and psycholinguistic data in the study of phonological and phonetic encoding: A parallel investigation of syllable frequency effects in brain-damaged and healthy speakers. *Language, Cognition and Neuroscience* 29: 714–727.

Perry, Conrad, Johannes C. Ziegler & Marco Zorzi. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology* 61: 106–151.

Perry, Conrad, Johannes C. Ziegler & Marco Zorzi. (2014). CDP++.Italian: Modelling sublexical and supralexical inconsistency in a shallow orthography. *PloS one* 9: e94291.

Primativo, Silvia, Pasquale Rinaldi, Shaunna O'Brien, Despina Paizi, Lisa S. Arduino & Cristina Burani. (2013). Bilingual vocabulary size and lexical reading in Italian. *Acta Psychologica* 144: 554–562.

Ševa, Nava, Padraic Monaghan & Joanne Arciuli. (2009). Stressing what is important: Orthographic cues and lexical stress assignment. *Journal of Neurolinguistics* 22: 237–249.

Spinelli, Giacomo, Simone Sulpizio & Cristina Burani. (2016). Q2Stress: A database for multiple cues to stress assignment in Italian. *Behavior Research Methods*, 1–14. Advanced online publication. doi: 10.3758/s13428-016-0845-7

Stella, Vitantonio & Remo Job. (2001). The PD/DPSS syllables. A database on syllable frequency in written Italian. *Giornale Italiano di Psicologia* 28: 633–639. Retrieved from http://www.dpss.unipd.it/materiali-e-strumenti-di-ricerca

Sulpizio, Simone, Lisa S. Arduino, Despina Paizi & Cristina Burani. (2013). Stress assignment in reading Italian polysyllabic pseudowords. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39: 51–68.

Sulpizio, Simone & Lucia Colombo. (2013). Lexical stress, frequency, and stress neighbourhood effects in the early stages of Italian reading development. *The Quarterly Journal of Experimental Psychology* 66: 2073–2084.

Sulpizio, Simone & Remo Job. (2010). Effect of syllable frequency in speech production and visual word recognition: evidence from Italian. *Giornale Italiano di Psicologia* 37: 707–718.

Sulpizio, Simone & Remo Job. (2013). Syllable frequency and stress priming interact in reading Italian aloud. In Markus Knauff, Michael Pauen, Natalie Sebanz & Ipke Wachsmuth (eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 1402–1407. Austin, TX: Cognitive Science Society.

Sulpizio, Simone & James M. McQueen. (2012). Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language* 66: 177–193.

Sulpizio, Simone, Giacomo Spinelli & Cristina Burani. (2015). Stress affects articulation planning in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance* 41: 453–461.

Thornton, Anna M., Claudio Iacobini & Cristina Burani. (1997). *BDVBD. Una base di dati sul vocabolario di base della lingua italiana* [BDVDB: A database for the Italian basic dictionary]. Roma: Bulzoni.

Vio, Claudio & Maria L. Tretti. (2008). Il trattamento del disturbo della lettura. Evidenze dell'efficacia di un intervento di automatizzazione della decodifica attraverso il software abilitativo Occhio alla lettera [Treating dyslexia: Evidence regarding the effectiveness of a decoding automatization treatment using Occhio alla lettera ('Watch the letter') software]. *Dislessia* 12: 289–300.

Authors' addresses

Simone Sulpizio
Faculty of Psychology
Vita-Salute San Raffaele University, Milan
Corso Bettini, 31
38068 Rovereto (TN)
Italy

simone.sulpizio@unitn.it

Giacomo Spinelli
Department of Psychology
University of Western Ontario
London, Ontario
Canada

gspinel@uwo.ca

Cristina Burani
Institute of Cognitive Sciences and Technology, Department of Life Sciences
University of Trieste
Via San Martino della Battaglia 44
00185 Rome
Italy

cristina.burani@istc.cnr.it